# CENTRAL-LINK-DEGENERACY IN PROTEIN CODING

By D. C. Peaslee*

[*Manuscript received June* 12, 1963]

*Summary*

In this paper it is proposed to interpret protein coding on a model allowing degeneracy in the central link of a triplet code: namely, that in addition to the "strong" links Au, Ua and Cg, Gc, some "weak" pyrimidine–pyrimidine links like Uu, Cc or Uc, Cu can also be functional.

From this postulate and the published data on artificial stimulation of polypeptide formation, a set of codes is deduced for the 20 amino acids. This code is in good agreement with others obtained by different lines of reasoning that generally place high weight on nitrous acid replacement data and natural substitution data, which are used here only in a secondary role. This agreement suggests that the codes obtained from current data are relatively free of bias introduced by particular lines of argument.

The novel features of the present treatment are most apparent when the codes are considered in terms of the sRNA ("acceptor") triplets rather than those of the mRNA ("donor"). Here the number of independent codes is greatly reduced in the present scheme; the price paid for this simplicity is the increased incidence of ambiguity in the code itself, without reference to possible suppressor mechanisms *in vivo*.

The total number of different acceptor triplets required by present data appears to be only 24, a number smaller than currently assumed. In reconciliation it may be noted that acceptors of different molecular weight for the same acid need *not* display different acceptor triplets.

## I. Introduction

It is now clear that a triplet code with considerable degeneracy governs the synthesis of protein by mRNA. The most detailed codes so far proposed rest on (*a*) measured stimulation by artificial polynucleotide mixtures of incorporation of amino acids into polypeptide chains *in vitro*; and (*b*) interpretation of observed amino acid replacements as one-base substitutions in the code. The codes so deduced turn out to display (*c*) a high degree of central-link-degeneracy: that is, multiple codes for the same amino acid in which the only variation occurs in the central link of the code triplet.

For some time it has seemed to the author that something like (*c*) is a rather attractive *a priori* hypothesis on simple mechanical grounds, while (*b*) is open to the objection that the experimental situation surrounding the measurements is often not under very good control. The present paper therefore reverses the logical order: a code is deduced on the basis of (*a*) plus (*c*) and is then applied to check (*b*). The

* Department of Theoretical Physics, Australian National University, Canberra.

codes obtained by these two different arguments are very similar, suggesting the conclusions that:

(i) both (*b*) and (*c*) have a high degree of validity (of the order of 90%);
(ii) the codes deduced from current data (*a*) are fairly objective in depending rather little on the particular line of deduction employed.

The main difference of the present approach from previous ones lies in the suggestion that a single acceptor may respond to more than one code instead of insisting on a one-to-one correspondence. This notion has two embarrassing consequences:

(1) As well as allowing redundancy (two codes for one amino acid), it also entails a fair amount of ambiguity* (two acids for one code) which is biologically unacceptable and can be explained away only by postulating suppressor or avoidance mechanisms *in vivo* that fail for *in vitro* measurements.
(2) It reduces to 24 the minimum number of acceptors necessary to interpret all current data, a number that seems rather too small.

Although these objections may ultimately prove decisive, we venture to present the following in the hope that the discussion will emphasize how important it is now to determine exactly the *acceptor* triplets on the sRNA as distinct from the code or *donor* triplets on the mRNA. Throughout this discussion mRNA stands for "messenger" RNA, which is copied directly from the DNA; and sRNA stands for "soluble" RNA, which effects the transfer of information from mRNA to the amino acid chain.

The present considerations raise an intriguing biological question (C. I. Davern, personal communication) without suggesting any answer. In spite of its inherent bias towards ambiguity, central-link-degeneracy can in principle permit the construction of totally unambiguous codes for up to 32 amino acids. Failure of this to occur in nature suggests the existence of other evolutionary criteria for an optimal code in addition to that of minimum ambiguity.

## II. Definitions

Bases in the mRNA triplets will be denoted by capital letters $X$, $Y$, $Z$, and bases in the sRNA triplets by lower case $x$, $y$, $z$; the triplets align so that $Xx$, $Yy$, and $Zz$ form pairs. We shall try to investigate the following hypothesis:

the end links must be $Xx$ and $Zz$ = Cg, Au, Ua, or Gc;
the central link can be $Yy$ = Cg, Au, Ua, Gc (strong); or
$Yy$ = pyrimidine–pyrimidine (weak).

The arguments for this hypothesis are simply geometrical. For a code read along a linear chain it seems essential that the end links of any element be quite firm to minimize abortions; but given firm end links, great additional strength in the central link may not be required—it must only be non-disruptive. This is why purine–purine central links can immediately be ruled out because of the greater size of the purines. A pyrimidine–pyrimidine link would contain a gap; but there might be weak binding across the gap, especially if the chains were slightly flexible. A look at the specific

* The only ambiguity considered here is in attachment of sRNA to mRNA; ambiguity in the enzyme linkage of sRNA to amino acid is ignored.

binding centers on these bases indicates that the most likely binding would be for Uc or Cu, with Uu and Cc less likely. We shall not assume any *a priori* preference among these pyrimidine–pyrimidine possibilities, trying instead to determine them from the data. This form of central-link-degeneracy is shown in Table 1. The immediate

TABLE 1

CENTRAL-LINK-DEGENERACY FOR MESSENGER ($Y$) AND ACCEPTOR ($y$) TRIPLETS

Strong codes are denoted by $s$, possible weak ones by $w$, although not all the latter may be realized in practice. Redundancy occurs among different codes in the same row, ambiguity among those in the same column

| $y$ \ $Y$ | U | C | A | G | |
|---|---|---|---|---|---|
| u | $(w)$ | $(w)$ | $s$ | | ← Redundancy |
| c | $(w)$ | $(w)$ | | $s$ | ← Redundancy |
| a | $s$ | | | | |
| g | | $s$ | | | |

↑ Ambiguity

question raised is the apparent equal likelihood of ambiguity and redundancy. On this point a few comments may be in order:

(1) If not all codes are necessarily significant, then it is possible to make a selection containing more redundancy than ambiguity: e.g. multiple acceptors for a given amino acid.

(2) The influence of possible ambiguity on viable systems will depend on competition between strong and weak codes. Presumably this will result in much less observable ambiguity than can be forced on artificial systems.

(3) The best current indication of central-link-degeneracy from an *in vitro* experiment does indeed show redundancy: ambiguity in a ratio of 1 : 1, as anticipated.

## III. CODING PATTERNS

### (a) *Pattern for* (X.Z)/(x.z) = $(U.U)/(a.a)$

Our analysis begins by considering an experiment (Weisblum, Benzer, and Holley 1962) that shows artificial coding of Leu by two acceptors: Leu I, coded by UC, not by U or UG; Leu II, coded strongly by UG, weakly by U, not by UC. Here Leu II seems to afford an ideal example of central-link-degeneracy, with $(X.Z)/(x.z)$ = (U.U)/(a.a) and $y$ = c; then $Y$ = G is a strong code, $Y$ = U is a weak code, and $Y$ = C does not code. This indicates a distinction between pyrimidine–pyrimidine links: Uc and hence presumably Cu function as weak central links, but Cc does not. No evidence on Uu bonds is available from this system. The weak code UUU for

Leu II is ambiguous with the well-known (Nirenberg and Matthaei 1961) strong code for Phe. The situation is shown in tabular form below:

| y \ Y | U | C | A | G |
|---|---|---|---|---|
| u | | | | |
| c | (Leu II) Phe | | | Leu II |
| a | | | | |
| g | | | | |

This tabulation suggests a systematic procedure for assigning other code triplets: namely, to construct similar patterns for the remaining 15 combinations $(X.Z)/(x.z)$. These combinations will all be independent if left–right differences are significant, as seems most probable. As an exercise, we shall attempt to carry out this programme as far as possible solely on the basis of artificial coding data. Variations in these data will be interpreted qualitatively but without insistence on exact coding ratios, as the intermediate mechanisms are not well understood.

(*b*) *Pattern for* (X.Z)/(x.z) = $(A.A)/(u.u)$

A strong code for Lys appears established (Gardner *et al.* 1962) as AAA. The corresponding acceptor is uuu, for which ACA should code weakly by the hypothesis above. Exactly this is suggested by Table 3 of Jones and Nirenberg (1962), where coding by various A/C ratios is reported. Coding by pure A is not included in this table, but among available entries the incorporation is a maximum for A/C = 2/1; it is comparably large for A/C = 4/1.

This illustrates the value of systematically varying the polynucleotide ratio in a single set of consistent experiments. Even though the quantitative values cannot be precisely interpreted, the qualitative pattern may be immediately significant. For example, the question of a weak Uu link could presumably be tested by a similar study of Lys incorporation with varying ratios of U/A. Such data do not seem to be currently available.

Now let us look for an ambiguity associated with the weak Lys code: that is, some other amino acid *strongly* coded by ACA, having $y$ = g. There are just three acids (Jones and Nirenberg 1962) besides Lys and Pro (code CCC) that respond strongly to an A/C mixture, and all of them appear to be $A_2C$: Asp*, Glu*, and Thr (here Asp* stands for Asp *or* $AspNH_2$; likewise for Glu*). But in fact (Gardner *et al.* 1962) Asp and Glu do *not* respond to A/C, while $AspNH_2$ and $GluNH_2$ appear to have codes $A_2C$. Assuming distinct codes, we must have ($AspNH_2$, $GluNH_2$, Thr) = (AAC, CAA, ACA), although not necessarily in that order.

Under our present form of central-link-degeneracy the strong codes AAC and CAA have respective weak codes ACC and CCA associated with them, while the strong code ACA has no associated weak code. Therefore, the pattern of amino acid incorporation as a function of A/C ratio should be more sharply centered around

A/C = 2/1 for one of these acids (the ACA code) than for the other two. Inspection of Table 3 of Jones and Nirenberg (1962) reveals Asp* as the candidate for ACA on this basis. We thus arrive at the following:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (?) | (Lys) | Lys | |
| c | | | | |
| a | | | | |
| g | | $AspNH_2$ | | |

As a by-product we obtain the pattern for $(X.Z)/(x.z)$ = (A.C)/(u.g)+(C.A)/(g.u) as set out below:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (?) | (Thr+$GluNH_2$) | Thr+$GluNH_2$ | |
| c | | | | |
| a | | | | |
| g | | His | | |

Here His has been added as a strong ACC or CCA code, since it seems to have no other (Gardner *et al.* 1962; Jones and Nirenberg 1962). Note that incorporation data do not generally permit distinction between right and left code orders.

### (c) *Pattern for* (X.Z)/(x.z) = *(C.C)/(g.g)*

It is known (Wahba *et al.* 1963) that Pro is coded by CCC, but $C_2A$ is also indicated (Gardner *et al.* 1962); again, variation of incorporation with C/A ratio (Jones and Nirenberg 1962) suggests both codes. We thus infer $y$ = u, $Y$ = A (strong), $Y$ = C (weak). In this case the older data (Matthaei *et al.* 1962; Speyer *et al.* 1962) definitely suggested $C_2U$ as a code; this is our first tentative indication that Uu also functions as a weak central link. These relations are indicated below:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (Pro?) | (Pro) | Pro | |
| c | | | | |
| a | Pro? | | | |
| g | | Pro? | | |

Ambiguously related strong codes in this tabulation would require $y$ = a or g, corresponding to codes $C_2U$ or $C_3$. The only possible candidate apparent in the references quoted would be a second acceptor for Pro. This would be a case in which potential ambiguity is suppressed by accidental identity of the amino acid for two ambiguously related acceptors.

(*d*) *Pattern for* (X.Z)/(x.z) = (*G.G*)/(*c.c*)

Codes for Gly have been reported as $G_2U$ (Matthaei *et al.* 1962; Speyer *et al.* 1962), $G_2A$ (Gardner *et al.* 1962), and GC (Wahba *et al.* 1963). Jones and Nirenberg (1962) obtained erratic and generally negative results for Gly, which suggests some failure in their system for this particular acid. The latter two codes for Gly correspond to $y$ = u; the code $G_2U$ then provides a second indication of Uu as a weak central link, which we now accept.

An ambiguously related strong code would have $y$ = a (code GUG) or $y$ = c (code GGG). The latter code apparently does not exist; the only candidates for GUG in addition to Gly are Cys (Matthaei *et al.* 1962) and Try (Matthaei *et al.* 1962; Speyer *et al.* 1962). But Cys is *not* (Wahba *et al.* 1962) GUG; and if Try were GGU or UGG, it should code weakly as $U_2G$, which was not reported (Matthaei *et al.* 1962; Speyer *et al.* 1962). We thus arrive at the following:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (Gly) | (Gly) | Gly | |
| c | | | | |
| a | Try | | | |
| g | | | | |

(*e*) *Pattern for* (X.Z)/(x.z) = (*U.C*)/(*a.g*)+(*C.U*)/(*g.a*)

In addition to Pro, only two acids (Matthaei *et al.* 1962; Speyer *et al.* 1962) give a positive response to UC: Leu I and Ser, and both appear to be $U_2C$. This must be acceptor Ser I, because Ser also codes by ACG but not by AC, AG, or CG (Gardner *et al.* 1962; Jones and Nirenberg 1962; Wahba *et al.* 1963).

It does not appear possible to state definitely that $y$ = a (no redundancy) for both Leu I and Ser I; we assume this for simplicity in the absence of contrary data.

(*f*) *Pattern for* (X.Z)/(x.z) = (*C.G*)/(*g.c*)+(*G.C*)/(*c.g*)

Two conspicuous candidates (Jones and Nirenberg 1962; Wahba *et al.* 1963) for a $C_2G$ code are Ala and Arg; but Arg also appears to respond (Gardner *et al.* 1962; Jones and Nirenberg 1962) to $A_2G$. In this case we must suppose two Arg acceptors, and take Ala and Arg I to be $C_2G$.

As in the preceding case, one cannot be sure on present evidence that $y$ = g (no redundancy) for both acceptors, but may assume it provisionally for simplicity.

(*g*) *Pattern for* (X.Z)/(x.z) = $(A.G)/(u.c)+(G.A)/(c.u)$

Both Glu and Arg II appear (Gardner *et al.* 1962; Jones and Nirenberg 1962) to be $A_2G$ but not (Matthaei *et al.* 1962; Speyer *et al.* 1962) $A_2U$, although AGU is also a possibility (Matthaei *et al.* 1962; Speyer *et al.* 1962) for Glu. The top row of the following tabulation includes these possibilities:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (Glu+Arg II) | (Glu+Arg II) | Glu+Arg II | |
| c | | | | |
| a | Met+Asp I | | | |
| g | | Ser II+Asp II | | |

The prevalence of secondary codes (II) suggests that the assignments here may be of reduced reliability.

To this we have added Met and Ser II as being uniquely AUG and ACG in all these references. Furthermore, it appears (Speyer *et al.* 1962) that Asp codes with AGU but not with UG, UA, UAC, or UCG; and also (Wahba *et al.* 1963) that it codes with ACG but not with CG. On this slender evidence we have had to assign Asp I and Asp II.

(*h*) *Remaining Patterns*

We have now made assignments for all but four acids: Cys, Ileu, Tyr, and Val. These four do *not* (Jones and Nirenberg 1962; Wahba *et al.* 1963) code with ACG and so must contain U in their codes. The two sets of unassigned patterns that remain are

$$(X.Z)/(x.z) = (\mathrm{U.G})/(\mathrm{a.c})+(\mathrm{G.U})/(\mathrm{c.a})$$

and

$$(X.Z)/(x.z) = (\mathrm{A.U})/(\mathrm{u.a})+(\mathrm{U.A})/(\mathrm{a.u}).$$

It is immediately tempting to associate these patterns with the remaining acids, which proves trivially easy.

Only the code $U_2G$ is reported (Matthaei *et al.* 1962; Speyer *et al.* 1962) for Val; and the code for Cys is known (Wahba *et al.* 1962) to have G on one end and U on the other, with the central member either G or U. There is a slight indication (Matthaei *et al.* 1962) that both these codes may be effective, as postulated below.

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | | | | |
| c | (Cys) | | | Cys |
| a | Val | | | |
| g | | | | |

The code for Ileu has been quoted both as $AU_2$ (Matthaei *et al.* 1962; Speyer *et al.* 1962) and as $A_2U$ (Gardner *et al.* 1962). It is known (Wahba *et al.* 1962) that the code for Tyr has A on one end, U on the other, and A or U in the middle; the older measurements (Matthaei *et al.* 1962; Speyer *et al.* 1962) indicate only $AU_2$, and there is no indication of ACU as a code. We therefore arrive at:

| *y* \ *Y* | U | C | A | G |
|---|---|---|---|---|
| u | (Ileu) | (Ileu) | Ileu | |
| c | | | | |
| a | Tyr | | | |
| g | | | | |

## IV. Comparison with Nitrous Acid Replacement Data and Natural Substitution Data

### (*a*) *Nitrous Acid Replacement Data*

A code has now been completely assigned so far as letter content is concerned, but no left–right order has been determined. The code is at least consistent in the sense that no letter assignments are duplicated more than once, so that left–right distinctions can make all codes unique. We can assign an arbitrary order for one asymmetric code and determine all other orders relative to this standard. For this purpose only replacement data are presently available; thus, in checking our letter assignments against replacement data, we obtain clarification on relative order.

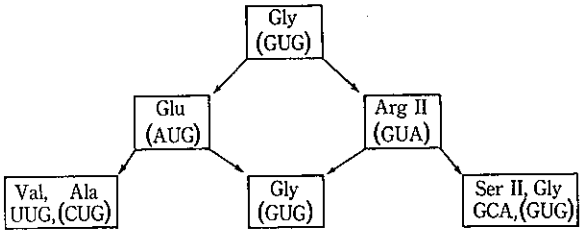


Fig. 1.—Replacement data interpreted by present codes. The order for Glu is an arbitrary standard.

The most direct replacement data (Henning and Yanofsky 1962) concern the substitutions Gly → Glu → Gly, Val, Ala, and Gly → Arg → Gly, Ser, Gly. Assignments consistent with one-base substitution and the codes above are given in Figure 1.

Table 5 of Speyer *et al.* (1962) gives a summary, with references, of nitrous acid replacement data. In addition to some substitutions already considered above there appear:

| (i) | (ii) | (iii) | (iv) |
|---|---|---|---|
| Ileu ↔ Val | Thr ↔ Ileu | Phe ↔ Leu | Pro ↔ Leu |
| | Thr ↔ Met | Phe ↔ Ser | Pro ↔ Ser |
| | Thr ↔ Ser | Phe ↔ Tyr | Ser ↔ Leu |

Here we have shown all transitions as reversible whether actually observed to be or not; transitions with undifferentiated Asp* or Glu* have been omitted. These replacements compare with the codes above as follows:

(1) Our conventional order in Figure 1 for Val is UUG; the Ileu code accessible by a one-base substitution is (UUA) with additional codes (UCA) and UAA.

(2) With this order established for Ileu, one has (CUA), (CCA), and CAA as the corresponding one-base substitutions leading to Thr. From Thr to Met implies GUA for Met, and Thr to Ser is consistent with GCA for Ser II as in Figure 1. Note that these assignments imply Ileu ↔ Met, which was not contained in Table 5 of Speyer *et al.* (1962), but has since been inferred by the same techniques (Jukes 1962).

(3) The code UUU for Phe allows all these transitions to be interpreted at once as one-base substitutions (Speyer *et al.* 1962). Both Leu I and Leu II can be involved, but only Ser I. Because of the symmetry of the Phe code, no right–left order can be assigned for the others.

(4) One needs only to add the weak code (CUC) for Pro to connect it by one-base substitutions with Leu I + Ser I = UUC + CUU. A one-base transition between Leu and Ser can only be between Leu II (UUU) and Ser I.

Additional nitrous acid replacement data are quoted by Jukes (1962), distinguishing between Asp and $AspNH_2$, and Glu and $GluNH_2$. The transitions of Glu are the same as those given earlier in this section. In addition the following are cited:

| | | |
|---|---|---|
| $AspNH_2$ ↔ Arg II | : | ACA ↔ (GCA) |
| $AspNH_2$ ↔ Lys | : | ACA ↔ AAA |
| $AspNH_2$ ↔ Ser II | : | ACA ↔ GCA |
| $GluNH_2$ ↔ Val | : | (AUC) × UUG |
| Asp ↔ Gly | : | AUG ↔ (GUG) |
| Asp ↔ Lys | : | AUG ↔ (AUA) |

Interpretation on the present codes is given at the right. The $GluNH_2$ ↔ Val transition is not possible with one-base substitution; this difficulty would vanish if the reported transition were a misidentified Glu ↔ Val. The transitions for Asp are interpreted for Asp I; they work equally for Asp II if the central U is replaced by C throughout.

### (*b*) *Natural Substitution Data*

These are mostly assembled from haemoglobin studies and are quoted by Jukes (1962). In addition to the above, they include:

| | | |
|---|---|---|
| Ala ↔ Val | : | (CUG) ↔ UUG |
| $AspNH_2$ ↔ Thr | : | ACA ↔ (CCA) |
| Glu ↔ Lys | : | AAG ↔ AAA |
| His ↔ Arg | : | CCA ↔ (GCA) |
| His ↔ Tyr | : | CCA × UUA |
| Glu ↔ $GluNH_2$ | : | AAG ↔ AAC |

The transition Ala ↔ Val suggests that the code CCG for Ala is not strong but weak and is accompanied by the weak code (CUG) and the strong one CAG. The transition His ↔ Tyr does not represent a single-base change on the present code; again, it is

conceivable that at least one out of all the transitions quoted may involve double replacement.

## V. Discussion

The central-link-degeneracy model is more concisely discussed in terms of acceptor triplets rather than codes. The patterns assembled above do not reveal any choice of $(x.z)$ for which both $y = c$ and $y = u$ occur. The biological significance of this is to eliminate the *a priori* possibility of triple ambiguity; such elimination would be expected as a natural result of evolution. It seems worth while to state this result explicitly:

$y = c$ and $y = u$ do not both occur for any choice of $(x.z)$.

This means that the total number of possible acceptor triplets is reduced to 48.

For each combination of $x.z$ the four choices of $y$ can be classified as follows:

$y = u$, triple redundancy possible ($Y = A, C, U$);
$y = c$, double redundancy possible ($Y = G, U$);
$y = a, g$, double ambiguity possible (with $y = u$, or $y = c$).

Table 2
Acceptor triplets

| $x.z$ | $y$ | Acceptor | $x.z$ | $y$ | Acceptor |
|---|---|---|---|---|---|
| a.a | a | Phe | c.g | ? | Arg I |
| | c | Leu II | g.c | u | Ala |
| c.c | a | Try | c.u | a | Met |
| | u | Gly | | g | Ser II |
| g.g | u | Pro | | u | Arg II |
| u.u | g | $AspNH_2$ | u.c | a | Asp I |
| | u | Lys | | g | Asp II |
| a.c | a | Val | | u | Glu |
| c.a | c | Cys | g.u | g | His |
| a.g, g.a | ? | Leu I, Ser I | | u | Thr |
| | | | u.g | u | $GluNH_2$ |
| a.u | u | Ileu | | | |
| u.a | a | Tyr | | | |

The degree of ambiguity still remaining in the codes is surprisingly high, and one must assume suppressor mechanisms to act *in vivo* but to fail for *in vitro* experiments, which are the source of practically all coding data.

The analysis above leads to the 24 acceptor triplets listed in Table 2, which is half the maximum number possible without triple ambiguity. That these triplets include all $(x.z)$ combinations and can account for 29 out of 31 observed amino acid replacements might suggest that not many more acceptors exist. This is a little at variance with the indications from direct measurements (e.g. Sueoka and Yamane 1962) of an average of the order of two acceptors per acid.

The distribution of bases among mRNA codes corresponding to Table 2 is U : C : A : G = 25 : 23 : 30 : 22, which is not significantly different from a uniform

ratio. For the end links of the acceptors ($x.z$), the observed ratio is u : c : a : g = 31 : 29 : 21 : 19; but the most striking deviation from uniformity is for the central acceptor link $y$, with the statistics u : c : a : g = 42 : 10 : 29 : 19. The bias here tends to maximize redundancy; it would be reduced if more acceptors were established, so that redundancy could more often be ascribed to multiple acceptors.

Comparison of the present code with that of Jukes (1963) and Wahba *et al.* (1963) is given in Table 3. The measure of agreement is very satisfactory, especially in view

TABLE 3
COMPARISON OF CODES

| Amino Acid | Codes from: | |
|---|---|---|
| | Jukes (1963); Wahba *et al.* (1963) | Present Study |
| Ala | CUG, CAG, CCG | CAG (CCG, CUG) |
| Arg | GUC, GCC, GAA | I GCC (+?)<br>II GAA (GCA, GUA) |
| Asp | GUA, GCA | I AUG<br>II ACG |
| $AspNH_2$ | UAA, CUA, CAA | ACA |
| Cys | GUU* | GGU (GUU) |
| Glu | AUG, AAG | AAG (ACG, AUG)* |
| $GluNH_2$ | AAC, AGG | AAC (ACC, AUC) |
| Gly | GUG, GAG, GCG | GAG (GCG, GUG) |
| His | AUC, ACC | CCA |
| Ileu | UUA, AAU | UAA (UCA, UUA) |
| Leu | UUC, UAU, UGU | I UCC/CCU (+?)<br>II UGU (UUU) |
| Lys | AAA, AUA | AAA (ACA, AUA) |
| Met | UGA | GUA |
| Phe | UUU | UUU |
| Pro | CUC, CCC, CAC | CAC (CCC, CUC) |
| Ser | CUU, ACG, UCC | I CCU/UCC (+?)<br>II GCA |
| Thr | UCA, ACA, CGC | CAA (CCA, CUA) |
| Try | UGG | GUG |
| Tyr | AUU* | AUU |
| Val | UUG | UUG |

* Indicates acid used to define conventional order among non-symmetric codes.

of the rather fragmentary experimental situation. To compare the codes of Table 3 in more detail, we group the acids into three classes:

Perfect agreement: Ala, Glu, Gly, Lys, Phe, Pro, Tyr, Val;

Good agreement: Arg, Asp, Cys, Leu, Met, Ser, Try;

Fair agreement: $AspNH_2$, $GluNH_2$, His, Ileu, Thr.

It is reassuring that these classes turn out to be in rough order of experimental accessibility.

Table 3 shows the agreement between previous and present codes; the differences are best seen by comparing acceptor triplets, as in Table 4. It is then apparent that the present assignment is a sort of minimal set; it is an experimental problem to determine whether the acceptors additional to this set actually occur. Significant differences in order for about one-third of the acceptors constitute a second type of discrepancy.

TABLE 4
COMPARISON OF ACCEPTORS

| Amino Acid | Codes from: Jukes (1963); Wahba *et al.* (1963) | Codes from: Present Study | Amino Acid | Codes from: Jukes (1963); Wahba *et al.* (1963) | Codes from: Present Study |
|---|---|---|---|---|---|
| Ala | gac, guc, ggc | guc | Leu | aag, aua, aca | aca, a?g |
| Arg | cag, cgg, cuu | c?g, cuu | Lys | uuu, uau | uuu |
| Asp | cau, cgu | uac, ugc | Met | acu | cau |
| $AspNH_2$ | auu, gau, guu | ugu | Phe | aaa | aaa |
| Cys | caa | cca | Pro | gag, ggg, gug | gug |
| Glu | uac, uuc | uuc | Ser | gaa, ugc, agg | cgu, g?a |
| $GluNH_2$ | uug, ucc | uug | Thr | agu, ugu, gcg | guu |
| Gly | cac, cuc, cgc | cuc | Try | acc | cac |
| His | uag, ugg | ggu | Tyr | uaa | uaa |
| Ileu | aau, auu | auu | Val | aac | aac |

## VI. ACKNOWLEDGMENTS

Without implying that they subscribe to any of the mistakes committed above, the author wishes to acknowledge helpful discussions with many people, particularly Drs. H. J. F. Cairns and C. I. Davern.

## ADDENDA

(1) Since this paper was written, a different scheme of central-link-degeneracy has been proposed by Eck (1963). One may note the following comparisons with the present study: (i) no physical basis is argued for the central-link-degeneracy hypothesis, so that some codes involve Gg or Aa links and seem rather unlikely; (ii) there is some mathematical rigidity in requiring just 32 codes, each one exactly twice degenerate, with no ambiguity; (iii) the importance is indicated of experimental study on *acceptor* rather than donor triplets. The same emphasis was obtained above and may fairly be considered a direct consequence of the central-link-degeneracy hypothesis.

(2) Professor S. Benzer (personal communication) has kindly informed the author that (i) the Leu II acceptor has been split experimentally into two components: Leu IIA coded by poly U, and Leu IIB coded by UG; and (ii) a single rabbit mRNA can code *all three* of these acceptors in a labelling experiment. The simplest modification of the scheme above is to assign acceptors aga, aua, and aca respectively to Leu I, Leu IIA, Leu IIB, and to call the mRNA code UCU. This implies that Cc is also a possible weak central link, contrary to the remarks above.

## VII. References


Eck, R. V. (1963).—*Science* **140**: 477.

Gardner, R. S., Wahba, A. J., Basilio, C., Miller, R. S., Lengyel, P., and Speyer, J. F. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 2087.

Henning, U., and Yanofsky, C. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 1497.

Jones, O. W. Jr., and Nirenberg, M. W. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 2115.

Jukes, T. H. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 1809.

Jukes, T. H. (1963).—*Biochem. Biophys. Res. Commun.* **10**: 155.

Matthaei, J. H., Jones, O. W., Martin, R. G., and Nirenberg, M. W. (1962).—*Proc. Nat Acad. Sci., Wash.* **48**: 666.

Nirenberg, M. W., and Matthaei, J. H. (1961).—*Proc. Nat. Acad. Sci., Wash.* **47**: 1588.

Speyer, J. F., Lengyel, P., Basilio, C., and Ochoa, S. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 441.

Sueoka, N., and Yamane, T. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 1454.

Wahba, A. J., Basilio, C., Speyer, J. F., Lengyel, P., Miller, R. S., and Ochoa, S. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 1683.

Wahba, A. J., Gardner, R. S., Basilio, C., Miller, R. S., Speyer, J. F., and Lengyel, P. (1962).—*Proc. Nat. Acad. Sci., Wash.* **49**: 116.

Weisblum, B., Benzer, S., and Holley, R. W. (1962).—*Proc. Nat. Acad. Sci., Wash.* **48**: 1449.